

# Treating entropy and conformational changes in implicit solvent simulations of small molecules

David L. Mobley<sup>¶\*</sup> and Ken A. Dill<sup>†</sup>

*Department of Pharmaceutical Chemistry,  
University of California, San Francisco, CA 94143*

John D. Chodera<sup>‡</sup>

*Department of Chemistry, Stanford University, CA 94305*

(Dated: August 10, 2007)

## Abstract

Implicit solvent models are increasingly popular for estimating aqueous solvation (hydration) free energies in molecular simulations and other applications. In many cases, parameters for these models are derived to reproduce experimental values for small molecule hydration free energies. Often, these hydration free energies are computed for a single solute conformation, neglecting solute conformational changes upon solvation. Here, we incorporate these effects using alchemical free energy methods. We find significant errors when hydration free energies are estimated using only a single solute conformation, even for relatively small, simple, rigid solutes. For example, we find conformational entropy ( $T\Delta S$ ) changes of up to 2.3 kcal/mol on hydration. Interestingly, these changes in conformational entropy correlate poorly ( $R^2 = 0.03$ ) with the number of rotatable bonds. The present study illustrates that implicit solvent modeling can be improved by eliminating the approximation that solutes are rigid.

Keywords: alchemical free energy calculation, implicit solvent, entropy, conformational change

---

\*Electronic address: [dmobley@gmail.com](mailto:dmobley@gmail.com)

†Electronic address: [dill@maxwell.compbio.ucsf.edu](mailto:dill@maxwell.compbio.ucsf.edu)

‡Electronic address: [jchodera@gmail.com](mailto:jchodera@gmail.com)

## I. INTRODUCTION

Solvation and desolvation processes drive many important biological and chemical processes, including binding, adsorption, protein folding, protein-protein interactions, and membrane formation. Hence, modeling the solvation component is important in computational biology and chemistry. However, because such computer simulations often involve large numbers of atoms and long time scales, it is valuable to compute accurate aqueous solvation (hydration) free energies as quickly as possible<sup>1-8</sup>. Because of this need, a common strategy is to use implicit solvent models, where water is treated as a continuum solvent, rather than in explicit molecular detail.

Perhaps the simplest and best-known such model is the Born model of ion solvation, which treats solvent as a high dielectric continuum and solves for the potential of a point charge in a spherical low dielectric cavity. It has proved relatively successful in estimating relative hydration free energies of ions with relatively few parameters<sup>9,10</sup> (ionic radii and internal and external dielectric constants). Other approaches include numerical solution of the Poisson – Boltzmann (PB) equation<sup>11-13</sup>, which describes the electrostatic interactions of charges embedded in a dielectric continuum, and Generalized Born (GB) models, which approximate the PB equation and generalize the Born equation to situations other than a single ion in a uniform dielectric<sup>14,15</sup>. A common feature of all of these approaches is that they require parameterization. At minimum, they require values for internal and external dielectrics and atomic radii. GB models often have additional parameters fit to maximize agreement with numerical solution of the PB equation on a test set, or with experimental solvation free energies.

Another class of solvation model is due to Cramer, Truhlar and others and is based on semi-empirical quantum mechanics on a solute in a dielectric continuum<sup>16,17</sup>. Such models are empirically optimized to reproduce solvation free energies<sup>16,17</sup>; these, too, require parameterization.

All of these solvation models have a number of adjustable parameters. Such parameters are obtained in different ways, including: (1) Optimizing the parameter set to minimize the discrepancy between hydration free energies computed with a single conformer and experimental hydration free energies (or transfer free energies between water and other solvents)<sup>10,16-22</sup>. (2) Optimizing GB models to reproduce single-conformer PB solvation free energies<sup>23-27</sup> (in some cases with additional adjustments). (3) Optimizing implicit solvent models to reproduce single-conformer hydration or charging free energies calculated from free energy calculations in explicit solvent<sup>28-30</sup>. (4) Optimizing implicit solvent models to reproduce forces and/or energies from explicit solvent

simulations for a fixed molecular geometry<sup>31</sup>.

In the first three of these methods, each hydration free energies is computed with a single solute conformation. In reality, a solute adopts an ensemble of conformations in both vacuum and solvent, and that ensemble can differ in the two states. Thus, using a single solute conformation is an approximation, and neglects conformational entropy and enthalpy changes of the solute. This may be especially important for approaches (1) and (2), above, where single conformation results are compared with experimental values. It is not obvious that single conformation hydration “free energies” (or *interaction free energies*, as we will call them here) should agree with experimental hydration free energies, which include any conformational enthalpic and entropic changes of the solute.

There appears to have been little work on conformational enthalpic and entropic changes in implicit solvent models. One study, on a few compounds, found that computed single-conformation solvation free energies agreed poorly with experiment for two compounds (which were previously believed to change conformation on solvation) until conformational changes were included<sup>10</sup>. Another study found that averaging over multiple conformations did not significantly affect computed water to octanol transfer free energies<sup>18</sup>, but did not test hydration free energies.

Here, our purpose is to learn the magnitude of errors incurred by fixing solutes into a single conformation when computing hydration free energies with implicit solvent models. We find that computed interaction free energies, on average, vary over a range of  $1.85 \pm 0.08$  kcal/mol depending on the solute conformation chosen, and differ substantially from true hydration free energies computed using the same parameters.

## II. THEORY

### A. Hydration free energies in implicit solvent

We consider the transfer of a single solute molecule from gas (with a 1M standard state) to water (1M standard state), at standard pressure and a particular temperature  $T$ . Following the notation of Deng and Roux<sup>32</sup>, we express the total potential energy  $U$  of the solute in water as

$$U(\mathbf{r}_u, \mathbf{r}_v) = U_u(\mathbf{r}_u) + U_v(\mathbf{r}_v) + U_{uv}(\mathbf{r}_u, \mathbf{r}_v) \quad (1)$$

where  $U$  is the potential energy, and  $u$  and  $v$  denote solute and solvent, respectively.  $\mathbf{r}_u$  denotes the full set of coordinates of all atoms in the solute, including the system's volume, and  $\mathbf{r}_v$  the similar quantity for the solvent. The classical hydration free energy, or excess chemical potential, is then given by:

$$\Delta G_{\text{hyd}} = -\beta^{-1} \ln \left[ \frac{\int \exp[-\beta U(\mathbf{r}_u, \mathbf{r}_v)] d\mathbf{r}_u d\mathbf{r}_v}{\int \exp[-\beta U_v(\mathbf{r}_v)] d\mathbf{r}_v \int \exp[-\beta U_u(\mathbf{r}_u)] d\mathbf{r}_u} \right] \quad (2)$$

The integrals run over the full simulation volume. Here  $\beta = 1/k_B T$ , where  $k_B$  is the Boltzmann constant and  $T$  is temperature. The numerator inside the logarithm is the partition function for the solute in water, and the denominator is the product of the partition functions for the solute in vacuum and for pure water. A key point here is that the hydration free energy involves integrals over all of the relevant (low potential energy) conformations of the solute and the water bath.

The principal approximation made in implicit solvent models is the replacement of the integral over the water degrees of freedom with a solute-water interaction free energy,  $G_{\text{int}}$ , effectively integrating out the solvent degrees of freedom. Then the solute experiences an effective potential energy (which already includes solvent entropic effects):

$$U_{\text{eff}}(\mathbf{r}_u) = U_u(\mathbf{r}_u) + G_{\text{int}}(\mathbf{r}_u) \quad (3)$$

where  $U_{\text{eff}}$  denotes an effective energy function that includes the solute potential energy and the interaction free energy ( $G_{\text{int}}$ ).  $G_{\text{int}}$  itself is the sum of two components:

$$\Delta G_{\text{int}} = \Delta G_{\text{polar}}(\mathbf{r}_u) + \Delta G_{\text{nonpolar}}(\mathbf{r}_u) \quad (4)$$

where  $G_{\text{polar}}$  treats the electrostatic interactions of the solute with the solvent (i.e. using continuum electrostatics) and  $G_{\text{nonpolar}}$  includes the nonpolar component of solvation. Then Eq. 2 reduces to:

$$\Delta G_{\text{hyd}} = -\frac{1}{\beta} \ln \left[ \frac{\int \exp[-\beta U_{\text{eff}}(\mathbf{r}_u)] d\mathbf{r}_u}{\int \exp[-\beta U_u(\mathbf{r}_u)] d\mathbf{r}_u} \right] \quad (5)$$

Here, the integrals over the solvent degrees of freedom are already incorporated into the effective potential energy  $U_{\text{eff}}(\mathbf{r}_u)$ .

Eq. 5 provides a rigorous way to compute hydration free energies within the framework of an implicit solvent model, including, in principle, solute conformational enthalpic and entropic changes. Essentially, it represents the free energy of turning on the implicit solvent term in an equilibrium ensemble, including any changes to the ensembles sampled as the solvent is turned on.

## B. Interaction free energies

It is common to estimate hydration free energies using single solute conformations<sup>10,16,18–22</sup> (perhaps for speed, or because conformation is assumed to be unimportant). Because these approaches keep the solute fixed in a single conformation for the transfer process, these are approximate. We refer to this quantity here as the *interaction free energy* and reserve the term *hydration free energy* for free energies computed including proper sampling of solute degrees of freedom.

It follows from Eq. 5 and Eq. 3 that the interaction free energy (Eq. 4) will be equal to the true hydration free energy in only two situations: (1) The solute prefers the same conformation in both environments, and this is the only relevant conformation, or (2)  $U_{\text{eff}}$  is independent of the solute conformation.

When computing the interaction free energy, a complication is how to choose which single conformation of the solute should be used<sup>58</sup>. We show below that different choices of solute conformations give different interaction free energy estimates of the hydration free energy, but there is only one correct thermodynamic estimate (using Eq. 5).

## C. Previous work

Alchemical free energy calculations are now commonly used with explicit solvent molecular mechanics models and forcefields to compute hydration free energies (for recent examples, see <sup>32–37</sup>). In the limit of sufficient sampling, these are equivalent to evaluating the ratio of partition functions in Eq. 2, for the given model and forcefield.

In contrast, as discussed above, implicit solvent simulations have more typically used only single conformations in estimating "hydration free energies". Several studies used different conformations in solvent and vacuum, but still neglected changes in solute entropy<sup>16,17</sup>. We are not aware of a single case using alchemical free energy methods for implicit solvent hydration free energies. Surprisingly, several studies comparing implicit solvent "hydration free energies" with those from explicit solvent also estimate free energies in this manner – the implicit solvent studies used single conformations, and the explicit solvent studies used alchemical methods<sup>21,22,38</sup>.

### III. THE SOLUTE TEST SET AND METHODS

#### A. Overview

We constructed a test set of 504 neutral small molecules with experimentally measured hydration free energies. We ran molecular dynamics simulations for each solute from a variety of different starting conformations, both in a GB water model, and in vacuum. We then estimated hydration free energies for these by analyzing the data using both alchemical free energy methods (which evaluate the ratio of partition functions as in Eq. 5), and by selecting single conformations and using Eq. 4, in the manner common in the literature.

#### B. Test set selection

We focus here only on neutral solute molecules, rather than on ions<sup>59</sup>. For our test set, we took as a starting point the experimental data compiled by Rizzo *et al.*<sup>19</sup> and removed the charged molecules. In an attempt to construct a set of all neutral compounds with known hydration free energies, we added (neutral) amino acid sidechain analogs and other compounds from our previous studies in explicit solvent<sup>22,35</sup>; and various small molecules from another set of 52 small molecules and associated references provided by J. Peter Guthrie<sup>39</sup>, removing redundancies. The result is a set containing 504 neutral small molecules.

The full set of experimental solvation free energies (1M gas to 1M water) and associated references are available in the Supporting Information. Most of these solute molecules are relatively small and fairly rigid, with an average of only 1.6 rotatable bonds per molecule, and an average molecular weight of 112 Da. Histograms of the number of rotatable bonds and molecular weight are shown in the Supporting Information.

#### C. Molecule preparation

Initial mol2 files for the small molecules in the Bordner *et al.* subset<sup>18</sup> of the Rizzo *et al.* set were obtained from the Supporting Information from that paper. The remaining mol2 files were generated from their chemical names using the tools `nam2mol`, `babel2`, and `Omega`, from OpenEye Scientific Software. To ensure consistency, conformers for the mol2 files from the Bordner set were generated from the mol2 files using `Omega`. All OpenEye tools used were those distributed

with version 1.5.0 of the OpenEye toolkits.

As a check of the Bordner mol2 files, we converted those mol2 files into chemical names and SMILES strings and compared with the desired chemical names and SMILES strings. We found that the mol2 files for iodoethane, iodomethane, and butanone actually contained slightly different molecules (iodopropane, iodoethane, and butan-2-one, respectively) so we generated new mol2 files containing the correct molecules.

In this work, protonation states were taken as the default generated by the OpenEye tools, and the protonation state was kept fixed throughout the simulations. It seems likely, however, that some of the molecules in this test set change protonation state upon solvation (acetic acid, for example, which has a pKa around 4.7<sup>40</sup>), and others may have multiple relevant protonation states in solvent. Since protonation states are not the focus of this work, we simply use the default protonation states, and keep these fixed throughout our calculations. This issue should probably be revisited in future work.

#### **D. Simulation setup**

For each molecule, partial charges were generated using the AM1-BCC method<sup>41,42</sup> in ANTECHAMBER v1.27 using the first molecular conformation generated by Omega.

Then, for each molecule, we generated starting conformations using Omega. We subsequently perform five independent sets of simulations (each with a vacuum simulation and a water simulation) for each molecule, so we wanted independent starting conformations for each, when possible, to help assess convergence. When Omega generated fewer than five conformations, the conformations it generated were re-used for all sets of simulations.

Following the charge calculation and the generation of starting conformations, ANTECHAMBER was used to generate generalized Amber force field (GAFF)<sup>43</sup> parameters for the small molecules and set them up for simulation in the Amber molecular dynamics package<sup>44</sup>, version 9. For GB simulations, radii were set as recommended in the Amber manual (different implicit solvent models have different recommended settings). Radii and screening factors had to be added to the source code of tLeap for bromine and iodine; screening factors were set to the default. Radii for bromine and iodine for the "bondi", "mbondi", and "mbondi2" radii schemes were set to 1.85 and 1.98 Å respectively, as in other work<sup>19</sup>.

### E. Simulation protocols:

For each molecule, we ran 5 sets of simulations, with each set consisting of a separate GB and vacuum simulation. The GB model used was Amber’s IGB=5, from Onufriev *et al.*<sup>26</sup>, with no SA term used during dynamics, though this was later added using reweighting. Each of these calculations was 10 ns long, using 2 fs timestep and a 16 Å nonbonded cutoff, and default RGB-MAX. Temperature was regulated (to 300 K) using Langevin dynamics with a friction coefficient of 1/ps. Center of mass motion was removed every 100 timesteps, and the RESPA<sup>45</sup> algorithm was employed to evaluate long range interactions only every two timesteps. Trajectory snapshots were saved every 5000 steps (10 ps) for later reprocessing.

### F. Reprocessing and computation of free energy differences

The Bennett acceptance ratio (BAR) method provides a minimum uncertainty estimate of free energy differences between two thermodynamic states<sup>46,47</sup>. This method requires that potential energies of simulation snapshots from each state be evaluated in the other state. We obtained these potential energies by reprocessing stored snapshots using the SANDER module in reprocessing mode (as described in the manual, with *imin*=5, *maxcyc* = 1). We reprocessed our original GB simulations in GB and in vacuum, and then reprocessed the vacuum simulations in GB and in vacuum to obtain the potential energy differences (including the GB solvation component, when using GB) needed for BAR.

Alchemical free energy calculations in explicit solvent typically require multiple intermediate alchemical states to transition between the solvated and vacuum states. Here, however, we found that overlap between the solvated and vacuum ensembles of the solute was sufficient that the distributions of forward and reverse work overlapped well, and it was unnecessary to add intermediate states to obtain reasonable statistical uncertainties.

Our initial simulations were run without a surface area penalty term, so molecular surface areas for each stored snapshot were calculated using OpenEye’s ZAP, and a nonpolar term was added to the total energies in solvent prior to analysis, for the single-snapshot schemes. For hydration free energies, the nonpolar contribution to the hydration free energy was computed using exponential averaging (the Zwanzig relation)<sup>48</sup> of the nonpolar potential energies. The nonpolar term was evaluated as  $G_{nonpolar} = \gamma SA + \beta$ , as in Ref<sup>19</sup>, with  $\gamma = 0.00542 \text{ kcal}/(\text{mol } \text{\AA}^2)$  and  $\beta = 0.92$



kcal/mol. A variety of recent work has suggested that this traditional form of the nonpolar term can be improved by including some treatment of dispersion interactions<sup>49–52</sup>, but we have not tested these alternate treatments here as our focus is not on the accuracy of implicit solvent models, but on the method for computing hydration free energies given an implicit solvent model.

However, for comparison, we also reprocessed the GB simulations using two alternate GB models available in Amber: the model of Tsui and Case<sup>25</sup>, and the GBn model of Mongan *et al.*<sup>27</sup>. This allowed us to rigorously compute hydration free energies for these models, as well, using exponential averaging<sup>48</sup>, because phase space overlap was very good.

Also, we reprocessed our stored configurations in OpenEye’s PB solver, ZAP, and used exponential averaging to compute solvation free energies using PB. ZAP settings were as default, with default radii, and an inner dielectric of 1 was used with a grid spacing of 0.5 Å. The same nonpolar component was used. We have made stored trajectories and parameters (in Amber format) available online ([http://www.dillgroup.ucsf.edu/~dmobley/paper\\_exports/trajectories\\_implicit\\_solvent.tar.gz](http://www.dillgroup.ucsf.edu/~dmobley/paper_exports/trajectories_implicit_solvent.tar.gz)).

Reported interaction free energies (single-conformer hydration free energies), below, were computed by selecting a single snapshot from each set of simulations for analysis. For each computed free energy, the overall value was taken as the average over the five separate sets of simulations.

In addition to computing the hydration free energy, we also computed the effective enthalpy and entropy of solvation, using  $T\Delta S = \Delta H - \Delta G$ , where  $\Delta H = \langle U_{GB} \rangle - \langle U_{vac} \rangle$ , where  $S$ ,  $H$ , and  $G$  are the entropy, enthalpy, and free energy, respectively, and  $U_{GB}$  is the potential energy in Generalized-Born solvent, while  $U_{vac}$  is the potential energy in vacuum. Since  $U_{GB}$  is an effective potential energy that already implicitly includes solvent entropies (as in Eq. 3), the entropy of solvation computed here only includes changes in *solute* entropy. Solvent entropy changes, since they are built in to the energy function, get included in the effective enthalpy of solvation. Thus, the effective enthalpy of solvation includes the solvent terms of Eq. 4, plus an additional term that includes the solute enthalpy change on solvation, which will be especially important in the case of conformational change.

## G. Error analysis

Statistical uncertainties were estimated for each simulation set using the asymptotic variance of the BAR method<sup>47</sup>, and Taylor expansion-based error propagation for exponential averaging. Standard error propagation was then used to compute uncertainties in the overall free energies.

For comparison, we took the standard error in the mean of the estimated  $\Delta G_{hyd}$  over the five sets of simulations. Because these were started from different conformations, this was in some cases larger than the estimate from the approach discussed above, from BAR. To be conservative, reported uncertainties below are the larger of these two estimated uncertainties.

## IV. RESULTS

### A. Comparison of hydration free energies with single conformation interaction free energies

Our main goal here is to compare different methods of estimating hydration free energies from implicit solvent models. To do so, we used the GB model of Onufriev *et al.*<sup>26</sup>. As our gold standard, we computed hydration free energies using extensive molecular dynamics sampling of the vacuum and solvated phases, and the Bennett acceptance ratio. Thus, these results include any solute conformational enthalpic and entropic changes (as in Eq. 5). We compare these results to interaction free energies computed using single solute conformations (Eq. 4).

We tested several different choices of solute conformations for computing interaction free energies: (1) The lowest potential energy conformation of the solute in solvent (the "BestSolv" scheme); (2) the lowest potential energy conformation of the solute in vacuum (the "BestVac" scheme); (3) the conformation of the solute in solvent that results in the most favorable interaction free energy (the "BestG" scheme); and (4) the conformation of the solute in solvent that results in the least favorable interaction free energy (the "WorstG" scheme). The BestSolv and BestG schemes are distinct, because internal energies are included in the potential energy when selecting the conformation for the BestSolv estimate, but neglected in selecting the conformation for the BestG estimate (since internal energies always cancel out when computing interaction free energies). Thus, BestG conformations can have larger internal strain energies.

The BestG and WorstG schemes span the range of interaction free energies. We find that the average value for this range is  $1.85 \pm 0.08$  kcal/mol across the whole test set of 504 small molecules. The distribution of ranges is shown in Figure 2. There are 314 small molecules with

ranges larger than 1 kcal/mol; 110 with ranges larger than 2.5 kcal/mol; 31 with ranges larger than 5 kcal/mol, and 5 with ranges larger than 7.5 kcal/mol. This indicates that, for many molecules, interaction free energies are quite sensitive to the conformation chosen. Furthermore, 187 small molecules have a range of values that is at least twice as large as the error relative to experiment – so the errors introduced by using single conformations are *larger* than those introduced by the force field in many cases.

We find that RMS differences between the computed hydration free energies and the interaction free energies can be up to 1.8 kcal/mol (Table I, Figure 1), depending on how the fixed solute conformation is chosen for interaction free energies. The full hydration free energies included both solute conformational enthalpic changes and entropic effects, so it was of interest to separate out the effects of entropic and enthalpic changes. We used the expression  $\Delta G = \Delta U_{eff} - T\Delta S$ , where  $U_{eff}$  is the effective energy function discussed above and implicitly includes solvent entropy changes. Thus,  $\Delta S$  here only includes solute entropic changes. We refer to  $\Delta U_{eff}$  as the effective enthalpy of hydration; it includes solute conformational enthalpic changes.

We compared effective enthalpies of hydration with computed interaction free energies (Table II). We find that solute conformational enthalpic changes contribute more to the overall error than do solute entropic changes. The average  $T\Delta S$  of hydration over the entire test set is only  $0.065 \pm 0.001$  kcal/mol.

However, we cannot conclude that solute entropic changes are unimportant. We find that 17 of the small molecules have  $T\Delta S$  of hydration larger than 0.5 kcal/mol (Fig. 3, and a table of the 17 in Supporting Information). Interestingly, it appears hard to predict which molecules will have large entropic changes on solvation: We find that there is essentially no correlation of  $T\Delta S$  with the number of rotatable bonds ( $R^2 = 0.03 \pm 0.01$ ) (in agreement with recent work from the Gilson laboratory<sup>53</sup>); correlations with the number of polar hydrogens and hydroxyl groups are also quite poor (data not shown).

Solute entropic changes are interesting, because they also end up being a strong predictor of the range of possible interaction free energies. In particular, the range of single-conformer hydration free energies correlates only very weakly with the number of rotatable bonds ( $R^2 = 0.05 \pm 0.01$ ) and the number of polar hydrogens ( $R^2 = 0.18 \pm 0.07$ ), but more strongly with  $T\Delta S$  of hydration ( $R^2 = 0.53 \pm 0.20$ ). We suspect this is because molecules which undergo conformational changes on hydration are often those which form strong internal electrostatic interactions in vacuum (for example, internal hydrogen bonding) which are often less favorable in solvent. In addition to

these conformational changes, these molecules often have very low entropy in their vacuum conformations due to the strength of these internal electrostatic interactions, and so entropy gains on hydration are particularly large.

It is interesting to note that of the single conformation approaches, the approach that uses conformations that are low potential energy in vacuum gave the lowest errors relative to full hydration free energies (Table I). Presumably, this is because the choice of conformation is less important in solvent than in vacuum due to roughness of the energy landscape.

Figure 4 shows examples of some of the changes in preferred conformation we observe on hydration, along with the resulting interaction free energies. At least with these parameters, acetic acid undergoes a conformational transition upon solvation that seems fairly straightforward: In vacuum, the molecule prefers to put positive charge near negative charge and reduce the total potential energy, but in solvent, atoms can interact with their image charges in solvent, making the alternate conformation more favorable; electrostatic interactions are also less strong in solvent<sup>60</sup>. Methyl formate is very similar; for both of these molecules, the BestG and BestSolv conformations are similar to one another, as are the BestVac and WorstG conformations. 2-methoxyphenol also shows a conformational change on solvation – but in this case, only the BestG conformation has the hydroxyl hydrogen pointing out (Fig. 4(c), right). This latter conformation is *extremely* unfavorable in vacuum due to the proximity of the two negatively charged oxygens. Since the interaction free energy is computed as the difference between the potential energy of the conformation in vacuum and the effective energy in solvent, molecules can achieve very favorable interaction free energies with solvent by adopting conformations that are very unfavorable in vacuum (and thus minimizing their apparent affinity for vacuum). This, of course, is purely an artifact of using single conformations to estimate hydration free energies, and is unphysical, resulting in a range of computed interaction free energies of 6.4 kcal/mol for methyl formate with these conformations.

Bis-2-chloroethyl ether and phenyl trifluoroethyl ether are also shown in Figure 4. In vacuum, bis-2-chloroethyl ether minimizes electrostatic clashes between the chlorine atoms, while in solvent, it can adopt other conformations. Phenyl trifluoroethyl ether appears similar, although it is less straightforward to interpret.

The key point, here, is that conformational changes can often occur on solvation, and when they occur, single-conformation "hydration free energies" (interaction free energies) are substantially incorrect, because they always incorrectly represent either the vacuum or solvent conformation. When using conformations appropriate for vacuum, these interaction free energies will tend to

underestimate the affinity for solvent (Figure 1, Table I), while the opposite happens when using conformations appropriate for solvent.

## B. Implications for parameterization

As noted, interaction free energies computed here using single conformations depend strongly on the solute conformation chosen. This dependence on conformation has significant implications for parameterization of implicit solvent models using single solute conformations. In particular, it raises an alarming possible scenario that is best illustrated by example. Suppose we selected the lowest potential energy conformation for each molecule in some existing solvent model (BestSolv scheme, here), computed experimental hydration free energies, and then were able to derive a new set of parameters (radii, etc.) that reproduced the experimental hydration free energies *exactly*. Then, we generated a new set of conformations for the same molecules using the scheme resulting in the most favorable hydration free energy (BestG scheme, here) and re-computed single-conformation hydration free energies; computed values would no longer agree completely with experiment. In fact, here, the RMS difference between the BestSolv and BestG schemes is 1.56 kcal/mol – with the same parameters! Given this, it seems reasonable to be skeptical of implicit solvent models which are parameterized using single conformations<sup>61</sup>. Given that the full hydration free energies reported here are relatively inexpensive to compute, these should probably be used in future parameterization efforts.

## C. Comparison with experiments

So far, we have compared interaction free energies computed using single solute conformations with hydration free energies computed using extensive sampling of solute conformations. This comparison tests only the importance of sampling, our main interest here, and does not test the underlying force field by comparing with experiment. Here, we provide a brief comparison with experimental results.

For Amber’s implicit solvent model IGB=5 (Onufriev *et al.*), we find that all of the approaches give relatively poor overall agreement with experiments (RMS errors from 2 to 2.4 kcal/mol), but still correlate well with experimental values (as observed previously on an overlapping test set<sup>19</sup>), so overall performance is mixed. Statistics are shown in Table III. It is interesting, however, to note

that the full free energy approach agrees substantially better with experiment (in terms of RMS error, mean error, and correlation coefficient) than either approach that selects good conformations from the solvent simulations. We discuss this further below. The free energy approach also results in roughly comparable agreement to the best of the other approaches ( $R^2 = 0.698 \pm 0.001$  with the full free energy approach, versus  $R^2 = 0.691 \pm 0.002$  with the best of the other approaches). Computed free energies are plotted versus experiment in Fig. 5.

We did not perform the same detailed comparison of different single-conformation schemes for the other implicit solvent models tested. However, we did compute hydration free energies using the full approach; statistics are shown in Table IV and results by compound, for the entire set, are available in the Supporting Information. Overall, the PB solver (OpenEye’s ZAP) performs marginally better than the GB methods, which is not too surprising, since these particular GB approaches (unlike many implicit solvent models) were parameterized (with some adjustments) to match PB results.

## V. DISCUSSION AND CONCLUSIONS

Our focus here is on how to parameterize implicit solvent models. In many cases, parameters are derived by selecting a single (presumably dominant) solute conformation and adjusting parameters to reproduce experimental vacuum-to-water transfer (hydration) free energies. Here, we performed molecular dynamics simulations of solutes in vacuum and water and computed hydration free energies using free energy methods (thus including entropic and conformational enthalpic changes). We then compared with results computed using single solute conformations, as a test of test the single conformation approximation.

We find that the single-conformation approaches lead to average errors between  $0.23 \pm 0.02$  and  $1.96 \pm 0.05$  kcal/mol (compared to the full free energy approach), depending on how the single solute conformation is chosen. These errors result from two factors. First, the solute samples an ensemble of conformations in each phase. Since the ensemble can change upon hydration, there can be solute entropic changes which are not included in the single conformation approaches. Second, the dominant conformation(s) of the solute can differ in the two phases. When this happens, no single conformation is appropriate for both phases. We find that the errors introduced by using single conformations are larger than those introduced by the force field in many cases. Moreover, errors introduced by using single conformations do not correlate strongly with the number of

rotatable bonds, although errors do correlate more strongly with changes in entropy on solvation.

In view of the results presented here, parameterizing single-conformation hydration free energy estimates to match experimental hydration free energies seems unwise. However, in addition to the alternative discussed here (parameterizing full hydration free energy estimates to match experiment) there is another sensible option which has been tried by several groups: Parameterizing single-conformation hydration free energy estimates in implicit solvent to match single-conformation estimates from explicit solvent simulations<sup>28–30</sup>. In this case, both computed free energies neglect solute conformational change, so at least this is an internally consistent way to derive parameters, and bypasses the conformation dependence problems described here.

Overall, we demonstrated an approach for computing hydration free energies in implicit solvent using free energy methods. This approach includes changes in solute conformational entropy and enthalpy that have often been neglected in implicit solvent studies of the solvation of small molecules. While it does not greatly improve agreement between hydration free energy estimates and experiment, it does not degrade results, and at least provides a consistent way to treat solute conformational changes on solvation. We believe this approach should be used in future efforts to parameterize implicit solvent models.

## **VI. ACKNOWLEDGMENTS.**

We thank Vijay S. Pande and Edgar Luttmann (Stanford) for helpful discussions and comments on the manuscript. We also thank Nathan Baker (Washington University) and Jason Wagoner (Stanford) for comments on the manuscript. We appreciate the support of NIH grant R01 GM 64592. JDC acknowledges support from Vijay Pande (Stanford) through an NSF grant for cyber-infrastructure (NSF CHE-0535616).

## **VII. SUPPORTING INFORMATION AVAILABLE:**

Coordinate files (mol2) with AM1-BCC partial charges for the small molecules in the test set used here; list of computed values for each compound with each of the implicit solvent models, and for the model of Onufriev, Bashford, and Case, using different conformations and different analysis methods; experimental solvation free energies (1M vacuum to 1M water) and references for the molecules in the test set; histograms of molecular weight and number of rotatable bonds for the

test set; an alternative version of Fig. 1; and a table of the 17 small molecules with  $T\Delta S$  larger than 0.5 kcal/mol. This information is available free of charge via the Internet at <http://pubs.acs.org>.

- 
- [1] Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Proteins: Struct., Funct., and Genet.* **1999**, 34, 4-16.
- [2] Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. *Biochemistry* **2005**, 44, 2059-2071.
- [3] Camacho, C. J.; Ma, H.; Champ, P. C. *Proteins: Struct., Funct., Bioinf.* **2006**, 63, 868-877.
- [4] Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; III, C. L. B. *J. Med. Chem.* **2004**, 47, 3032-3047.
- [5] Huang, N.; Kalyanaraman, C.; Jacobson, M. P. *J. Chem. Inf. Model* **2006**, 46, 243-253.
- [6] Ozkan, S. B.; Wu, G. A.; Chodera, J. D.; Dill, K. A. *Proc. Nat. Acad. Sci. USA* **2007**, In press, <http://www.pnas.org/cgi/doi/10.1073/pnas.0703700104>.
- [7] Zagrovic, B.; Sorin, E. J.; Pande, V. S. *J. Mol. Biol.* **2001**, 313, 151-169.
- [8] Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B. *Biopolymers* **2003**, 68, 91-109.
- [9] Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, 89, 5588-5593.
- [10] Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, 98, 1978-1988.
- [11] Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, 157, 671-679.
- [12] Zauhar, R. J.; Morgan, R. S. *J. Mol. Biol.* **1985**, 186, 815-820.
- [13] Gilson, M. K.; Honig, B. H. *Prot.: Struct., Funct., and Genet.* **1988**, 3, 32-52.
- [14] Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, 51, 129-152.
- [15] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **2000**, 90, 6127-6129.
- [16] Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, 113, 8305-8311.
- [17] Jason D. Thompson, C. J. C.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, 108, 6532 - 6542.
- [18] Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. *J. Phys. Chem. B* **2002**, 106, 11009-11015.
- [19] Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. *J. Chem. Theory Comput.* **2006**, 2, 128-139.
- [20] Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, 95, 5610-5620.
- [21] Mohan, V.; Davis, M. E.; McCammon, J. A.; Pettitt, B. M. *J. Phys. Chem.* **1992**, 96, 6428-6431.
- [22] Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Pande, V. S. **2007**, Submitted.
- [23] Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, 102, 9571-9576.
- [24] Dominy, B. N.; III, C. L. B. *J. Phys. Chem. B* **1999**, 103, 3765-3773.
- [25] Tsui, V.; Case, D. A. *Biopolymers (Nucl. Acid. Sci.)* **2001**, 56, 275-291.



- [26] Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383-394.
- [27] Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theor. Comput.* **2007**, *3*, 156 - 169.
- [28] Zhang, L. Y.; Gallicchio, E.; Friesner, R. E.; Levy, R. M. *J. Comp. Chem.* **2001**, *22*, 591-607.
- [29] Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B.* **1997**, *101*, 5239-5248.
- [30] Swanson, J. M. J.; Adcock, S. A.; McCammon, J. A. *J. Chem. Theor. Comput.* **2005**, *1*, 484-493.
- [31] Swanson, J. M. J.; Wagoner, J. A.; Baker, N. A.; McCammon, J. A. *J. Chem. Theor. Comput.* **2007**, *3*, 170-183.
- [32] Deng, Y.; Roux, B. *J. Phys. Chem. B* **2004**, *108*, 16567-16576.
- [33] Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740-5761.
- [34] Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.
- [35] Mobley, D. L.; Dumont, È.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242-2254.
- [36] Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616-17626.
- [37] Xu, Z.; Luo, H. H.; Tieleman, D. P. *J. Comp. Chem.* **2007**, *28*, 689-697.
- [38] Jean-Charles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 1454-1455.
- [39] Guthrie, J. P. Manuscript in preparation.
- [40] Fasman, G. D. *Handbook of Biochemistry and Molecular Biology*; CRC Press: 2000 Corporate Blvd., Boca Raton, Florida, 33431, 1977.
- [41] Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132-146.
- [42] Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623-1641.
- [43] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
- [44] Case, D. A.; III, T. E. C.; Darden, T.; Gohlke, H.; Luo, R.; K. M. Merz, J.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668 - 1688.
- [45] Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990-2001.
- [46] Bennett, C. H. *J. Comp. Phys.* **1976**, *22*, 245-268.
- [47] Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.
- [48] Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420-1426.
- [49] Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479-499.
- [50] Pitera, J. W.; van Gunsteren, W. F. *JACS* **2001**, *123*, 3163-3164.

- [51] Choudhury, N.; Pettit, B. M. *Mol. Sim.* **2005**, *31*, 457-463.
- [52] Wagoner, J. A.; Baker, N. A. *Proc. Nat. Acad. Sci. USA* **103**, 8331-8336.
- [53] en A. Chang, C.; Chen, W.; Gilson, M. K. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 1534.
- [54] Schurhammer, R.; Engler, E.; Wipff, G. *J. Phys. Chem. B* **2001**, *105*, 10700-10708.
- [55] Srinivasan, J.; III, T. E. C.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401-9409.
- [56] Tsui, V.; Case, D. A. *JACS* **2000**, *122*, 2489-2498.
- [57] “The PyMOL Molecular Graphics System”, 2002 <http://www.pymol.org>.
- [58] The assumption may be that the solute conformation is irrelevant.
- [59] Experimental hydration free energies for ions are really measured for ion pairs, and then separated into hydration free energies for individual ions using extrathermodynamic assumptions like the so-called TATB hypothesis, which assumes that hydration free energies of two isosteric large hydrophobic ions of opposite sign are equal. Recent work has called this hypothesis, and related assumptions, into question<sup>54</sup>. At the very least, the issue of comparing with experimental ion hydration free energies is a complicated one, and one we choose to avoid here.
- [60] As noted in the Methods section, acetic acid is one of the molecules in this test set that is likely to change protonation state on hydration. In this study, protonation states were kept fixed, in keeping with common practices in parameterizing implicit solvent models.
- [61] In some sense, this consideration applies even to the GB models considered here<sup>25-27</sup>, since these GB models were derived from an original GB model which was parameterized to match PB calculations using PARSE radii as a starting point<sup>55,56</sup>. PARSE radii were themselves derived to fit experimental hydration free energies when using single solute conformations<sup>10</sup>.

Scheme	RMS error	Mean error
BestSolv	0.390+/-0.053	-0.123+/-0.016
BestVac	0.338+/-0.014	0.154+/-0.013
BestG	1.796+/-0.043	-1.126+/-0.062
WorstG	0.882+/-0.007	0.726+/-0.022

TABLE I: **RMS errors of interaction free energies relative to computed hydration free energies.** Single-conformer interaction free energies differ from computed hydration free energies. Shown are RMS and mean errors, relative to the computed hydration free energies, depending on the scheme for choosing the conformer for the calculation of interaction free energies.

Scheme	RMS error	Mean error
BestSolv	$0.519 \pm 0.053$	$-0.188 \pm 0.022$
BestVac	$0.231 \pm 0.022$	$0.089 \pm 0.009$
BestG	$1.960 \pm 0.049$	$-1.190 \pm 0.069$
WorstG	$0.789 \pm 0.006$	$0.661 \pm 0.019$

TABLE II: **RMS errors of interaction free energies relative to computed effective enthalpies of hydration.** Single-conformer interaction free energies differ from computed effective hydration enthalpies (which include solvent, but not solute, entropic effects). Shown are RMS and mean errors, relative to the computed effective hydration enthalpies, depending on the scheme for choosing the conformer for the calculation of interaction free energies.

<b>Approach</b>	<b>RMS error</b> (kcal/mol)	<b>Mean error</b> (kcal/mol)	<b>Correlation coefficient (<math>R^2</math>)</b>
$\Delta G_{hyd}$	$2.145 \pm 0.003$	$-0.654 \pm 0.091$	$0.698 \pm 0.001$
BestSolv	$2.287 \pm 0.017$	$-0.777 \pm 0.096$	$0.689 \pm 0.005$
BestVac	$2.055 \pm 0.006$	$-0.500 \pm 0.089$	$0.691 \pm 0.002$
BestG	$3.300 \pm 0.031$	$-1.780 \pm 0.124$	$0.671 \pm 0.004$
WorstG	$1.947 \pm 0.004$	$0.072 \pm 0.087$	$0.671 \pm 0.002$

TABLE III: **Agreement with experiment using different schemes.** Shown are the RMS and mean differences from experiment, and the correlation with experiment, for the implicit solvent hydration free energy calculations ( $\Delta G_{hyd}$ ) and the various single-conformation approaches.

Solvent model	RMS error (kcal/mol)	Mean error (kcal/mol)	Correlation coefficient ( $R^2$ )
IGB=5 <sup>26</sup>	2.145 $\pm$ 0.003	-0.654 $\pm$ 0.091	0.698 $\pm$ 0.001
PB (ZAP)	2.014 $\pm$ 0.008	-0.796 $\pm$ 0.082	0.774 $\pm$ 0.001
IGB=1 <sup>25</sup>	2.137 $\pm$ 0.006	-0.645 $\pm$ 0.091	0.737 $\pm$ 0.001
IGB=7 <sup>27</sup>	2.433 $\pm$ 0.002	-1.056 $\pm$ 0.098	0.685 $\pm$ 0.001

TABLE IV: **Agreement with experiment using different implicit solvent models.** Shown are the RMS and mean differences from experiment, and the correlation with experiment, for the implicit solvent hydration free energy calculations using various different implicit solvent models (the IGB=x models are different GB models in Amber), as described in the text.

FIG. 1: **Calculated hydration free energies, versus error with single-conformation schemes.** Calculated hydration free energies are plotted, versus the error from these values when using different choices of single-conformations for computing interaction free energies. The different schemes are: BestG, the conformation resulting in the most favorable interaction free energy; WorstG; the conformation resulting in the least favorable interaction free energy; BestVac, the conformation with the lowest potential energy in vacuum; BestSolv, the conformation with the lowest potential energy in solvent. The  $y = 0$  line indicates exact agreement between calculated interaction free energies and calculated hydration free energies. Values below the line overestimate the affinity for water; those above the line underestimate the affinity for water.

FIG. 2: **Distribution of range of single-conformation hydration free energies.** Single-conformation hydration free energies are sensitive to the choice of conformation used. For each compound, the range is the difference between the minimum and maximum single conformation hydration free energies obtained. Shown here is a histogram of the ranges for the 504 small molecules in the test set.

FIG. 3: **Number of rotatable bonds versus  $-T\Delta S$  of solvation.** The correlation between  $-T\Delta S$  and the number of rotatable bonds is only  $R^2 = 0.03$ , and there are 17 small molecules with  $-T\Delta S$  larger than 0.5 kcal/mol.

FIG. 4: **Sample conformations for molecules with significant ranges in computed hydration free energies.** For each molecule, the left shows the lowest potential energy conformation in vacuum (BestVac) scheme from one simulation. (a): The BestVac conformation (left) yields an interaction free energy of -8.95 kcal/mol; the BestSolv conformation (right) yields -17.75 kcal/mol. Worst and BestG conformations are similar to these two, respectively. (b): The BestVac conformation (left) yields -9.72 kcal/mol; the BestSolv conformation (right) yields -14.83 kcal/mol. WorstG and BestG conformations are similar to these two, respectively. (c) The BestVac conformation (left) yields -8.33 kcal/mol; the BestG conformation (right) yields -14.73 kcal/mol. BestSolv and WorstG conformations are similar to BestVac. (d): The BestVac conformation (left) yields -8.84 kcal/mol; the BestSolv conformation (right) yields -11.38 kcal/mol. WorstG and BestG conformations are similar to these two, respectively. (e): The BestVac conformation (left) yields -8.67 kcal/mol and the BestSolv conformation (right) yields -11.38 kcal/mol. The WorstG and BestG conformations are similar to these two, respectively. Images were made with PyMOL<sup>57</sup>.

FIG. 5: **Computed hydration free energies, and range of single-conformation interaction energies, versus experiment.** Computed hydration free energies are shown as black squares; vertical bars denote the range of interaction free energies possible depending on the choice of solute conformation. The diagonal line is the  $x = y$  line.

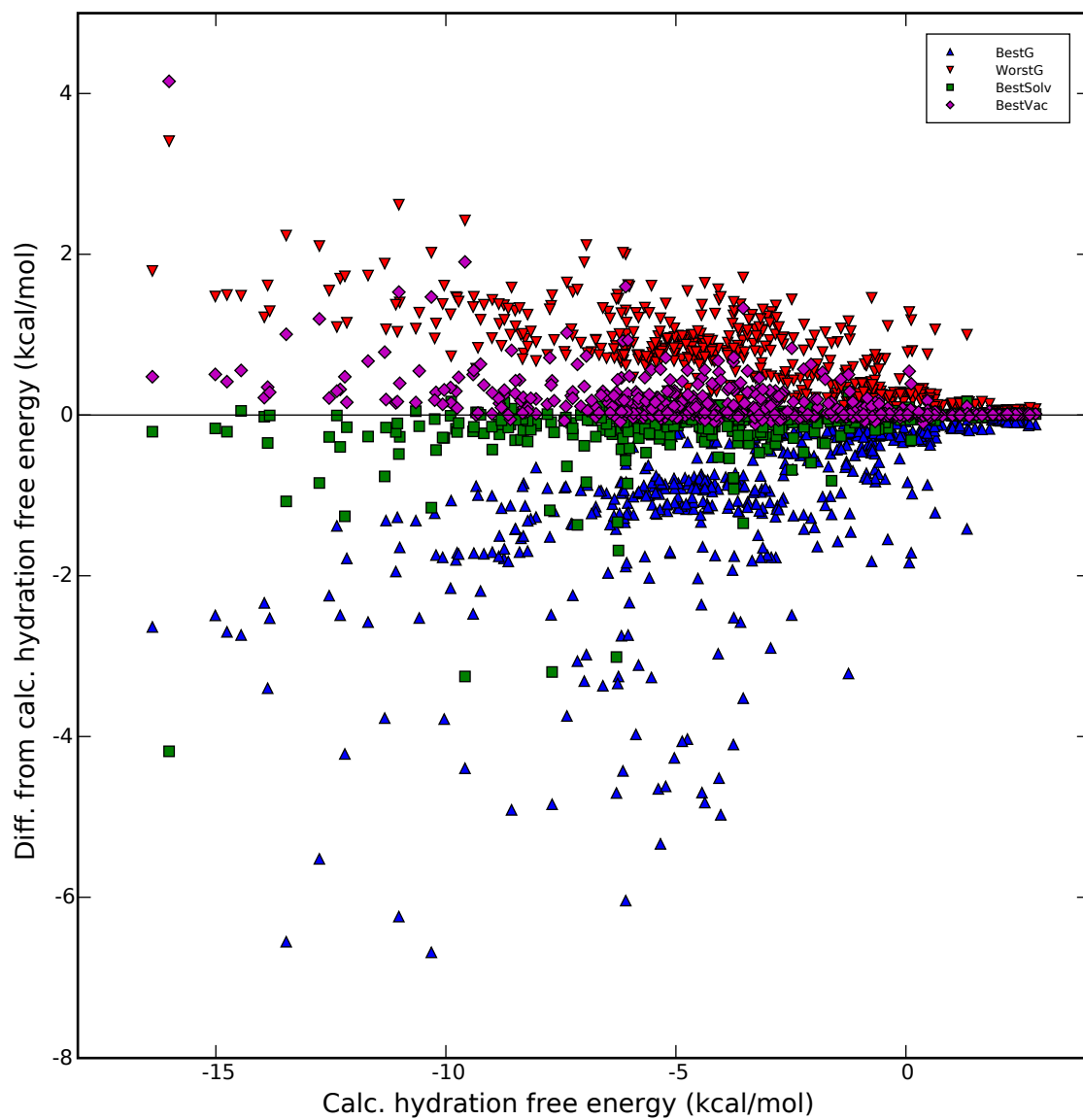


FIG. 1: Figure 1.



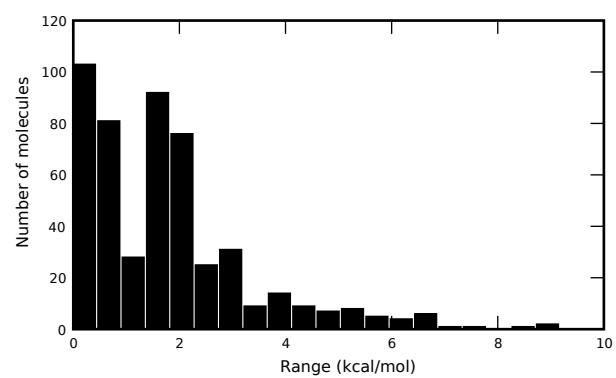


FIG. 2: Figure 2

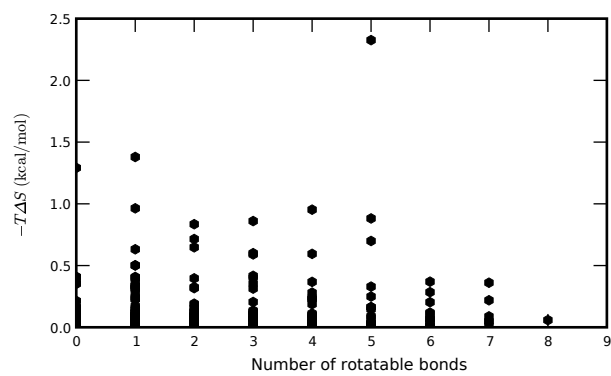
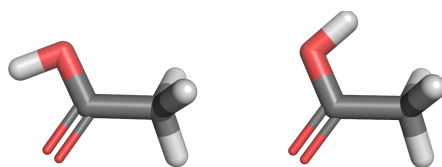
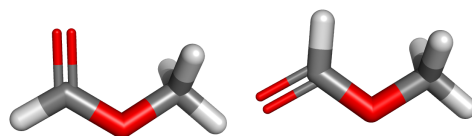


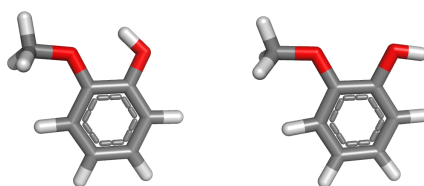
FIG. 3: Figure 3.



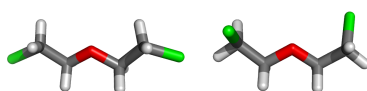
(a)acetic acid



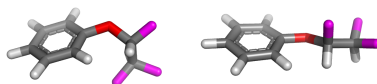
(b)methyl formate



(c)2-methoxyphenol



(d)bis-2-chloroethyl ether



(e)phenyl trifluorethyl ether

FIG. 4: Figure 4.

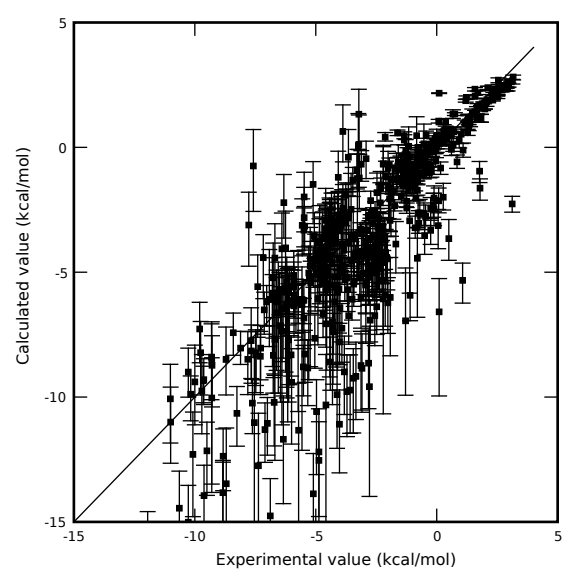


FIG. 5: Figure 5